

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-318938

(43)Date of publication of application : 16.11.2001

(54)Int.Cl. G06F 17/30
G06F 19/00
G06T 7/60
G09B 29/00

(21)Application number : 2000-135928

(71)Applicant : INTERNATL BUSINESS MACH
CORP <IBM>

(22)Date of filing : 09.05.2000

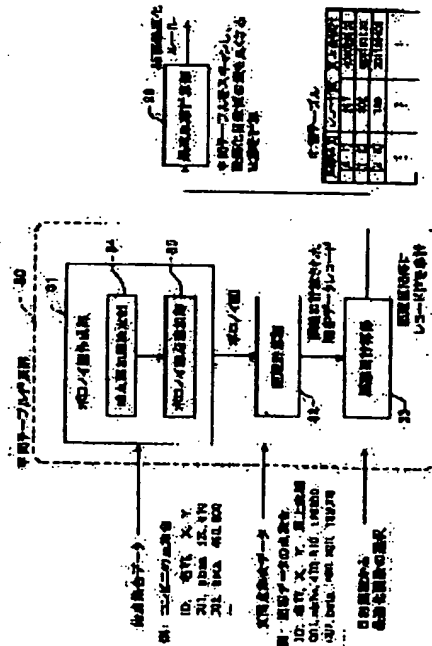
(72)Inventor : MORIMOTO YASUHIKO
KUBO HARUNOBU

(54) METHOD AND DEVICE FOR MINING SPACE DATA AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a space data mining method for finding out a distance itself and an azimuth itself for optimizing a certain purpose to be requested by many analytical operations without previously determining a distance and an azimuth and deriving a space correlation rule.

SOLUTION: The space data mining device for calculating an optimum distance from a data base including space information such as an address is provided with an input means for inputting an object function necessary for distance optimization, an intermediate table preparation part 30 for generating an intermediate table by calculating a distance between a start point and a question point on the basis of start point set data and question point set data stored in a data base, and an optimum distance calculation part 39 for calculating a distance for optimizing the value of the object function inputted by the input means on the basis of the intermediate table generated by the preparation part 30.



LEGAL STATUS

[Date of request for examination] 23.03.2001

[Date of sending the examiner's decision of rejection] 29.06.2004

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3659318

[Date of registration] 25.03.2005

[Number of appeal against examiner's decision] 2004-19608

【特許請求の範囲】

【請求項1】 住所などの空間情報を含むデータベースの中から空間的な法則を導き出す空間データマイニング方法であって、
前記データベースから始点または始点群を与えるステップと、

前記空間的な法則を導き出すために調べる目的関数を定義するステップと、

定義された前記目的関数を最適化するような前記始点または前記始点群からの距離区間を算出するステップとを含むことを特徴とする空間データマイニング方法。

【請求項2】 前記目的関数は、分析業務に要求される距離自体が与えられていない関数であることを特徴とする請求項1記載の空間データマイニング方法。

【請求項3】 前記始点または前記始点群の定義、および前記目的関数の定義を入力パラメータとして入力するステップとを更に含むことを特徴とする請求項1記載の空間データマイニング方法。

【請求項4】 前記距離区間を算出するステップは、前記始点群からなる始点集合データと定義された前記目的関数とに基づいて中間テーブルを生成すると共に、生成された当該中間テーブルに基づいて前記データベースにおける各質問点の属性値を距離の値に応じて集計することを特徴とする請求項1記載の空間データマイニング方法。

【請求項5】 算出された前記始点または前記始点群からの距離区間を地図上に表示するステップとを更に含むことを特徴とする請求項1記載の空間データマイニング方法。

【請求項6】 住所などの空間情報を含むデータベースの中から空間的な法則を導き出す空間データマイニング方法であって、
前記データベースから始点または始点群を与えるステップと、

与えられた前記始点または前記始点群から方位を定義するステップと、

前記空間的な法則を導き出すために調べる目的関数を定義するステップと、

定義された前記目的関数を最適化するような前記始点または前記始点群からの方位区間を算出するステップとを含むことを特徴とする空間データマイニング方法。

【請求項7】 前記目的関数は、分析業務に要求される方位自体が与えられていない関数であることを特徴とする請求項6記載の空間データマイニング方法。

【請求項8】 算出された前記方位区間は、前記目的関数を最適化する方位を数値として算出することを特徴とする請求項6記載の空間データマイニング方法。

【請求項9】 算出された前記方位区間は、検索対象となるデータの範囲として方位を算出するのに適した前記始点または前記始点群からの等距離範囲を選択すること

を特徴とする請求項6記載の空間データマイニング方法。

【請求項10】 算出された前記始点または前記始点群からの方位区間を地図上に表示するステップとを更に含むことを特徴とする請求項6記載の空間データマイニング方法。

【請求項11】 空間情報を含むデータベースの中から方位に関する空間的な法則を導き出すためのデータテーブルを生成する空間データマイニング方法であって、
前記データベースから始点の集合および質問点の集合を与えるステップと、

前記始点の集合と前記質問点の集合との距離の上限を指定するステップと、

前記質問点に対して前記始点との距離を計算するステップと、

計算された前記始点との距離が指定された前記上限に含まれる質問点に対して、当該始点との角度を計算するステップと、

計算された前記始点との角度を用いてデータテーブルを生成するステップとを含むことを特徴とする空間データマイニング方法。

【請求項12】 住所などの空間情報を含むデータベースの中から最適な距離を計算する空間データマイニング装置であって、

距離最適化に必要な目的関数を入力する入力手段と、

前記データベースにおける始点集合データおよび質問点集合データに基づいて始点と質問点との距離を計算して中間テーブルを生成する中間テーブル生成手段と、

前記中間テーブル生成手段によって生成された前記中間テーブルに基づいて前記入力手段により入力された前記目的関数の値を最適化する距離を計算する最適距離計算手段とを備えたことを特徴とする空間データマイニング装置。

【請求項13】 前記中間テーブル生成手段は、

前記データベースにおける始点集合データからボロノイ図を作成するボロノイ図作成手段と、

前記ボロノイ図作成手段によって作成された前記ボロノイ図と前記データベースにおける質問点集合データとに基づいて各始点と各質問点との距離を計算してデータレコードを生成する距離計算手段と、

調べたい目的関数から最適化関数を選択して距離区間毎に最適値に必要なレコード値を前記データレコードから集計する距離別計算手段とを含むことを特徴とする請求項12記載の空間データマイニング装置。

【請求項14】 前記ボロノイ図作成手段は、入力始点数に応じて平面四分割を繰り返し、前記入力始点数を末端の分割平面ピクセルに振り分けて各ピクセルの中にある始点の中から1つを当該ピクセルの代表点として選択し、各段ピクセルを中間ノードとする四分木を作成し、当該四分木の各ノードを広さ優先順に最上段から走査し

て順序付けされた始点集合を出力することを特徴とする請求項13記載の空間データマイニング装置。

【請求項15】 住所などの空間情報を含むデータベースの中から最適な方位を計算する空間データマイニング装置であって、
方位最適化に必要な目的関数を入力する入力手段と、
前記データベースにおける始点集合データおよび質問点集合データに基づいて、始点から特定方向を0度とする角度を用いて質問点の位置する方位を含む中間テーブルを生成する中間テーブル生成手段と、
前記中間テーブル生成手段によって生成された前記中間テーブルに基づいて前記入力手段により入力された前記目的関数の値を最適化する方位を計算する最適方位計算手段とを備えたことを特徴とする空間データマイニング装置。

【請求項16】 前記中間テーブル生成手段は、
前記データベースにおける始点集合データからボロノイ図を作成するボロノイ図作成手段と、
前記ボロノイ図作成手段によって作成された前記ボロノイ図と前記データベースにおける質問点集合データとに基づいて各始点と各質問点との距離を計算する距離計算手段と、
前記距離計算手段により計算された距離の中から、指定された上限距離以内の質問点の前記始点からの方位を計算して前記中間テーブルのレコードとする方位計算手段と、
調べたい目的関数から最適化関数を選択して方位区間毎に最適値に必要なレコード値を前記レコードから集計する方位別計算手段とを含むことを特徴とする請求項15記載の空間データマイニング装置。

【請求項17】 住所などの空間情報を含むデータベースの中から最適な距離または方位を計算して出力する空間データマイニング装置であって、
分析業務に要求される距離自体および方位自体が与えられていない目的関数を入力する入力手段と、
前記データベースにおける始点集合データおよび質問点集合データに基づいて始点と質問点との距離または方位を求めると共に、前記目的関数の値を最適化する最適距離または最適方位を計算する最適距離・方位計算手段と、
前記最適距離・方位計算手段により計算された最適距離または最適方位を地理情報システムの画面上に表示する表示手段とを備えたことを特徴とする空間データマイニング装置。

【請求項18】 前記表示手段は、前記最適距離・方位計算手段により計算された最適距離を各始点を中心とする円形領域にて表示することを特徴とする請求項17記載の空間データマイニング装置。

【請求項19】 前記表示手段は、前記最適距離・方位計算手段により計算された最適方位を各始点からの扇形

領域にて表示することを特徴とする請求項17記載の空間データマイニング装置。

【請求項20】 住所などの空間情報を含むデータベースの中から空間的な法則を導き出す空間データマイニング装置であって、
前記データベースから始点または始点群を与える始点付与手段と、
前記空間的な法則を導き出すために調べる目的関数を定義する目的関数定義手段と、

10 前記目的関数定義手段により定義された前記目的関数を最適化するような前記始点または前記始点群からの距離区間を算出する距離区間算出手段とを含むことを特徴とする空間データマイニング装置。

【請求項21】 住所などの空間情報を含むデータベースの中から空間的な法則を導き出す空間データマイニング装置であって、
前記データベースから始点または始点群を与える始点付与手段と、

20 前記始点付与手段により与えられた前記始点または前記始点群から方位を定義する方位定義手段と、
前記空間的な法則を導き出すために調べる目的関数を定義する目的関数定義手段と、
前記目的関数定義手段により定義された前記目的関数を最適化するような前記始点または前記始点群からの方位区間を算出する方位区間算出手段とを含むことを特徴とする空間データマイニング装置。

【請求項22】 空間情報を含むデータベースの中から方位に関する空間的な法則を導き出すためのデータテーブルを生成する空間データマイニング装置であって、
30 前記データベースから始点の集合および質問点の集合を与える始点/質問点付与手段と、
前記始点/質問点付与手段により付与された前記始点の集合と前記質問点の集合との距離の上限を指定する距離上限指定手段と、
前記質問点に対して前記始点との距離を計算する距離計算手段と、
40 前記距離計算手段により計算された前記始点との距離が前記距離上限指定手段により指定された前記上限に含まれる質問点に対して、当該始点との角度を計算する角度計算手段と、
前記角度計算手段により計算された前記始点との角度を用いてデータテーブルを生成するデータテーブル生成手段とを含むことを特徴とする空間データマイニング装置。

【請求項23】 距離自体および方位自体が与えられていない目的関数に基づいて住所などの空間情報を含むデータベースの中から空間的な法則を導き出すプログラムを格納した記憶媒体であって、
50 前記プログラムは、前記データベースから始点または始

点群を与えるステップと、与えられた前記始点または前記始点群からの距離または方位を定義するステップと、調べたい目的関数を定義するステップと、定義された前記目的関数を最適化するような、前記始点または前記始点群からの距離区間または方位区間を算出するステップとを含むことを特徴とする記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、空間データマイニングにおけるデータベース処理に関し、より詳しくは、空間データマイニングの基本機能である最適距離または最適方向を算出する手法、装置等に関する。

【0002】

【従来の技術】データベースにある住所などの空間情報を空間的な文脈で解釈し、その空間的な法則を大量のデータの中から導き出す新しい技術を空間データマイニングと呼ぶ。この空間データマイニングは、大量データに対して、コストの高い空間・幾何学演算を行なう必要があり、技術的にはかなり難しい問題に直面する場合が多い。そのために、空間データマイニングはあまり発展しているとは言えず、未開拓の研究分野である。しかしながら、その一方で、空間データマイニングは、情報産業の中でも最もビジネスボリュームの大きなデータベースやGIS (Geographical Information System: 地理情報システム) 分野を大いに発展させる可能性を秘めた基礎技術と考えられ、ビジネス的、また技術的にも大変有望な技術分野である。

【0003】ここで、既存の空間データマイニングシステムとして、あらかじめ距離を定めた上で空間相関ルールを導き出すものが知られている。例えば、J. Han 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

「close to 公園」→「住宅地である」(支持度5%、確信度80%)、

「地価下落」→「far from 駅」(支持度10%、確信度70%)

のような距離述語を含んだ空間相関ルールを導き出している。

【0004】また、既存の空間データマイニングシステムとして、あらかじめ方位を定めた上で空間相関ルールを導き出すものも知られている。例えば、上記J. Han 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900

「west of 公園」→「住宅地である」(支持度5%、確信度80%)、

「地価下落」→「north of 駅」(支持度10%、確信 50

度70%)

のような方位述語を含んだ空間相関ルールを導き出すことが可能である。

【0005】

【発明が解決しようとする課題】しかしながら、J. Han 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500

【0006】例えば、「A地区で、現金自動預払機(ATM)の単位距離あたりの存在密度を最大化する、コンビニエンスストアからの半径を求めよ」といった検索や、「強い大気汚染が広がっているのは、各ゴミ処分場からどの方向であるかを求めよ」といったような検索には、既存のデータマイニングシステムでは対応することができない。

【0007】本発明は、以上のような技術的課題を解決するためになされたものであって、その目的とするところは、予め距離や方位を定めた上で空間相関ルールを導き出すことなく、多くの分析業務で要求される距離自体、方位自体を求める技術を提供することにある。また他の目的は、距離自体、方位自体を求める際に、マイニング作業の高速化を図ることにある。更に他の目的は、ユーザ(お客様)に対して利便性の高いマイニング出力結果を提供することにある。

【0008】

【課題を解決するための手段】かかる目的のもと、本発明は、予め距離や方位を定めた上で空間相関ルールを導き出すのではなく、入力パラメータとして距離や方位の定義、始点集合の定義、目的関数の定義から、多くの分析業務にて要求される、ある目的を最適化する距離自体または方位自体を求める空間データマイニングを提供するものである。即ち、本発明は、住所などの空間情報を含むデータベースの中から空間的な法則を導き出す空間データマイニング方法であって、データベースから始点または始点群を与えるステップと、与えられた前記始点または前記始点群から距離または方位を定義するステップと、空間的な法則を導き出すために調べる目的関数を定義するステップと、定義された前記目的関数を最適化

7
 するような前記始点または前記始点群からの距離区間または方位区間を算出するステップとを含むことを特徴としている。

【0009】ここで、この目的関数は、分析業務に要求される距離自体または方位自体が与えられていない関数であることを特徴としている。また、距離の定義、始点または始点群の定義、および目的関数の定義を入力パラメータとして入力するステップとを更に含むことを特徴とすることができる。更に、距離区間を算出するステップは、始点群からなる始点集合データと定義された目的関数とに基づいて中間テーブルを生成すると共に、生成された中間テーブルに基づいてデータベースにおける各質問点の属性値を距離の値に応じて集計することとすれば、計算時間を大幅に短縮することができる点で好ましい。また、算出された始点または始点群からの距離区間または方位区間を地図上に表示するステップとを更に含むことを特徴とすれば、計算されたルールを視覚的に把握することが可能となり、ユーザに対する利便性を向上させることができる。

【0010】ここで、算出された方位区間は、この目的関数を最適化する方位を数値として算出することとすることができる。また、算出された方位区間は、検索対象となるデータの範囲として方位を算出するのに適した始点または始点群からの等距離範囲を選択することとすることができる。始点または始点群から無限遠まで計算するのは殆ど不可能であり、最適化される領域を定めることは有効である。

【0011】また、本発明は、空間情報を含むデータベースの中から方位に関する空間的な法則を導き出すためのデータテーブルを生成する空間データマイニング方法であって、データベースから始点の集合および質問点の集合を与えるステップと、始点の集合と質問点の集合との距離の上限を指定するステップと、質問点に対して始点との距離を計算するステップと、計算された始点との距離が指定された上限に含まれる質問点に対して、始点との角度を計算するステップと、計算された始点との角度を用いてデータテーブルを生成するステップとを含むことを特徴とすることができる。ここで、質問点とは、例えば顧客データの点集合等が挙げられ、実際に始点や始点群からの距離、方位が計算される対象である。

【0012】本発明を装置から捉えると、本発明は、住所などの空間情報を含むデータベースの中から最適な距離を計算する空間データマイニング装置であって、距離最適化に必要な目的関数を入力する入力手段と、データベースにおける始点集合データおよび質問点集合データに基づいて始点と質問点との距離を計算して中間テーブルを生成する中間テーブル生成手段と、この中間テーブル生成手段によって生成された中間テーブルに基づいて入力手段により入力された目的関数の値を最適化する距離を計算する最適距離計算手段とを備えたことを特徴と

することができる。

【0013】ここで、この中間テーブル生成手段は、データベースにおける始点集合データからボロノイ図を作成するボロノイ図作成手段と、このボロノイ図作成手段によって作成されたボロノイ図とデータベースにおける質問点集合データとに基づいて各始点と各質問点との距離を計算してデータレコードを生成する距離計算手段と、調べたい目的関数から最適化関数を選択して距離区間毎に最適値に必要なレコード値をデータレコードから集計する距離別計算手段とを含むことを特徴とすることができる。更に、このボロノイ図作成手段は、入力始点数に応じて平面四分割を繰り返し、入力始点数を末端の分割平面ピクセルに振り分けて各ピクセルの中にある始点の中から1つをピクセルの代表点として選択し、各段ピクセルを中間ノードとする四分木を作成し、この四分木の各ノードを広さ優先順に最上段から走査して順序付けされた始点集合を出力することとすれば、処理を高速化することが出来る点で好ましい。また、この四分木の構造を事前に計算しておき、メモリに蓄積するように構成すれば、この四分木の構造は距離最適化や方位最適化に関するマイニングで頻繁に利用することから、マイニング作業を更に高速化することができる点で好ましい。

【0014】一方、本発明は、住所などの空間情報を含むデータベースの中から最適な方位を計算する空間データマイニング装置であって、方位最適化に必要な目的関数を入力する入力手段と、このデータベースにおける始点集合データおよび質問点集合データに基づいて、始点から特定方向を0度とする角度を用いて質問点の位置する方位を含む中間テーブルを生成する中間テーブル生成手段と、この中間テーブル生成手段によって生成された中間テーブルに基づいて入力手段により入力された目的関数の値を最適化する方位を計算する最適方位計算手段とを備えたことを特徴とすることができる。ここで、この中間テーブル生成手段は、このデータベースにおける始点集合データからボロノイ図を作成するボロノイ図作成手段と、このボロノイ図作成手段によって作成されたボロノイ図とデータベースにおける質問点集合データとに基づいて各始点と各質問点との距離を計算する距離計算手段と、この距離計算手段により計算された距離の中から、指定された上限距離以内の質問点の始点からの方位を計算して中間テーブルのレコードとする方位計算手段と、調べたい目的関数から最適化関数を選択して方位区間毎に最適値に必要なレコード値をこのデータレコードから集計する方位別計算手段とを含むことを特徴とすることができる。

【0015】本発明を他の観点から捉えると、本発明は、住所などの空間情報を含むデータベースの中から最適な距離または方位を計算して出力する空間データマイニング装置であって、分析業務に要求される距離自体お

よび方位自体が与えられていない目的関数を入力する入力手段と、このデータベースにおける始点集合データおよび質問点集合データに基づいて始点と質問点との距離または方位を求めると共に、この目的関数の値を最適化する最適距離または最適方位を計算する最適距離・方位計算手段と、最適距離・方位計算手段により計算された最適距離または最適方位を地理情報システムの画面上に表示する表示手段とを備えたことを特徴とすることができる。

【0016】ここで、この表示手段は、最適距離・方位計算手段により計算された最適距離を各始点を中心とする円形領域にて表示することを特徴とすることができる。また、この表示手段は、最適距離・方位計算手段により計算された最適方位を各始点からの扇形領域にて表示することを特徴とすることができる。これらによれば、求めた最適距離や最適方位を地図上に解かり易く表示することが可能となり、お客様の使い勝手を各段に向上させることができる。

【0017】また、本発明は、住所などの空間情報を含むデータベースの中から空間的な法則を導き出す空間データマイニング装置であって、データベースから始点または始点群を与える始点付与手段、空間的な法則を導き出すために調べる目的関数を定義する目的関数定義手段、定義された目的関数を最適化するような始点または始点群からの距離区間を算出する距離区間算出手段、与えられた始点または始点群から方位を定義する方位定義手段、定義された目的関数を最適化するような始点または始点群からの方位区間を算出する方位区間算出手段とを含むことを特徴とすることができる。更に、データベースから始点の集合および質問点の集合を与える始点/質問点付与手段、付与された始点の集合と質問点の集合との距離の上限を指定する距離上限指定手段、質問点に対して始点との距離を計算する距離計算手段、計算された始点との距離が指定された上限に含まれる質問点に対して、始点との角度を計算する角度計算手段、計算された始点との角度を用いてデータテーブルを生成するデータテーブル生成手段とを含むことを特徴とすることができる。

【0018】また更に、本発明は、距離自体および方位自体が与えられていない目的関数に基づいて住所などの空間情報を含むデータベースの中から空間的な法則を導き出すプログラムを格納した記憶媒体であって、このプログラムは、データベースから始点または始点群を与えるステップと、与えられた始点または始点群からの距離または方位を定義するステップと、調べたい目的関数を定義するステップと、定義されたこの目的関数を最適化するような、始点または始点群からの距離区間または方位区間を算出するステップとを含むことを特徴とすることができる。この記憶媒体としては、例えばCD-ROM等の可搬性のある媒体の他、ネット等を介してプログ

ラムをダウンロードするための、プログラムを提供する側のハードディスク等の記憶媒体、ダウンロード等によってプログラムの提供を受けたユーザ側のハードディスク等の記憶媒体を含むものである。

【0019】

【発明の実施の形態】以下、添付図面に示す実施の形態に基づいてこの発明を詳細に説明する。まず、本実施の形態における空間データマイニングの理解を容易にするために、本方式のモデリングとアルゴリズムについて説明する。図1は、本実施の形態におけるモデリング1として、距離最適化エンジンによる出力例を示した図である。地図11上に、始点(または始点群)12~14であるコンビニエンスストア(CS)から、目的関数を最適化するような距離(あるいは距離区間)が、所定の半径を有する円15~17で表示される。図1に示す例では、「ひったくりの発生率を最大化するコンビニエンスストアからの距離を求めよ。」といった検索から、その目的関数を最適化するような始点12~14からの距離が示されている。

【0020】この出力例では、例えば、「ひったくり」→(「コンビニエンスストア」,「[0,100]」),

「5」件といった内容が出力される。その意味は、「ひったくり」の発生率を最大化するのは、「コンビニエンスストア」から半径「0」m以上「100」m以内で、半径(m)あたり「5」件発生している。である。始点からの距離の出力としては、以下のような例がある。

「ひったくり」→(「駅」,「[50,180]」),「6.1」件

「ひったくり」→(「銀行」,「[200,-]」),「2.2」件

「強盗」→(「銀行」,「[60,200]」),「1.3」件
「殺人」→(「飲食店」,「[0,50]」),「0.4」件

【0021】図2は、本実施の形態におけるモデリング2として、方位最適化エンジンによる出力例を示した図である。地図11上に、始点(または始点群)12~14であるコンビニエンスストア(CS)から、目的関数を最適化するような始点12~14からの方位区間18~20が算出されて示されている。図2に示す例では、「ひったくりの発生率を最大化するコンビニエンスストアからの方角を求めよ。」といった検索から扇型の領域で表現される角度が示されている。この出力例では、例えば、

「ひったくり」→(「コンビニエンスストア」,「[0,100]」),「5」件

といった内容が出力される。その意味は、「ひったくり」の発生率を最大化するのは、「コンビニエンスストア」から方位「0」°以上「100」°以内で、10°あたり「5」件発生している。である。始点からの方位の出力としては、その他に、

「ひったくり」→(「神社」,「[120,240]」),

「6.1」件

等の出力例が挙げられる。

【0022】図3は、本実施の形態における距離・方位最適化エンジンのアルゴリズムの概略を説明するためのフローチャートである。まず、始点(または始点群)を与える(ステップ101)。即ち、距離最適化の場合には距離を計算する際の起点となる地図上のエンティティ(実体)を指定し、方位を最適化するにはどの地点から見た方角なのかを計算するために方角基準点として地図上のエンティティを指定する。この始点は、各計算毎に1つでも複数でもよく、それを始点あるいは始点群(集合)とする。次に、その始点(または始点群)からの距離・方位を定義する(ステップ102)。方位では、どこを起点とすることを定義する。但し、常に距離をユークリッド距離として定める場合には、距離に対する定義は不要となる。そして、例えば、ある商品の売り上げ総額、犯罪の発生件数などの調べたいものの目的関数を定義する(ステップ103)。その後、その目的関数を最適化するような始点からの距離区間、あるいは方位区間を算出する(ステップ104)。その後、距離・方位のバケット集計と最適化距離の計算が実行される。尚、 n を質問点数、 m を始点点数とし、 $N=n+m$ と仮定すると、距離算出の平均実行時間は $O(N)$ で表わすことができる。即ち、質問点に対応する始点を見つけるのに、平均計算量で $O(\log n)$ 、距離の計算は $O(1)$ で可能であり、この処理に $O(n \log n)$ かかる。しかし、距離がユークリッド距離である場合は、後述する四分木構造(quaternary incremental method)を利用することで、平均実行時間を $O(n)$ とすることができる。ここで質問点とは、処理の対象となるデータベースにおける顧客データ等である。

【0023】図4は、データベース例を示した図である。本実施の形態では、図に示すようなスキーマ(仕様)を有するデータベースと関連付けされた統合化地理情報システムの存在を前提としている。このデータベースの各スキーマには、データを区別するためのID情報と、位置(座標)情報が含まれている。この位置情報は、住所データで表わされる場合の他、地図情報に対応した座標情報等が格納されている。また、下線で示される数値属性や、下線した斜体文字で示されるカテゴリカル属性が含まれている。このデータベースから、距離や方位に関する最適化ルールがマイニングされる。

【0024】図5は、入力パラメータとして始点群や基点群(方位の始点群)の定義例を示した図であり、図3のフローチャートのステップ101にて示した距離の始点群や方位の始点群を定義する例を示している。この始点や基点の定義としては、例えば、郵便局、学校、警察等のエンティティが指定される。ここで、例えば、郵便局(ALL)であれば、集配局や特定郵便局等の複数のカテゴリカル属性に基づく郵便局を統合して定義される。同様に学校(ALL)であれば、各種学校や小学校、中学校

等のカテゴリカル属性を統合して定義される。また、例えば、駅であれば、乗降客 X 以上や乗降客 X 未満等の数値属性に基づいて、始点群や基準群を定義できる。同様に、例えば、コンビニエンスストア(コンビニ)では、売上 X 以上、売上 X 未満等の数値属性に基づく定義が可能である。

【0025】図6は、距離や方位の定義例を示した図であり、図3のフローチャートにおけるステップ102の定義例である。これらの定義は、最適化ルールをマイニングする場合に、入力パラメータとして指定されるものである。距離の定義としては、ユークリッド距離やネットワーク距離がある。このユークリッド距離はボロノイ図などを利用して計算されるが、高速に計算する場合やごく近距離の場合には、地図の実体とかい離することがある。また、ネットワーク距離はダイクストラ法などで計算される。このダイクストラ法は、各ノードへの最短距離をスタートノードの周辺から1つずつ確定して徐々に範囲を広げていき、最終的に全てのノードへの最短距離を求める方法であり、計算時間は多く必要となるが地図上での実体を反映することができる。尚、本実施の形態では、ユークリッド距離の場合に適用される計算法を用いており、それ以外の距離では計算法が全く異なる。ユークリッド距離として決める場合には、距離の定義は不要である。また、方位の定義としては、例えば、北を 0° として時計回りに 360° で1周する方位スケールを定義する等があり、これらの方位スケール上で方位区間が定義される。

【0026】図7は、目的関数の定義例を示した図であり、図3のフローチャートにおけるステップ103の定義例である。これらの目的関数は最適化ルールをマイニングの際に指定されるものである。この目的関数の定義としては、各スキーマに対して、図の下線で示される数値(あるいは数値として派生する)属性、下線した斜体文字で示されるカテゴリカル(あるいはカテゴリ値として派生する)属性を用いて定義することができる。例えば、顧客スキーマとして、数値属性を用いて「支持率 S 以上の顧客の「平均年収」最大化距離」や、カテゴリカル属性を用いて「支持率 S 以上の顧客「年齢60歳以上」の顧客比率最大化距離」等の定義や、例えばATMスキーマとして「支持率 S 以上の顧客の「ATM数/顧客数」最大化距離」等を定義することができる。

【0027】次に、図3のフローチャートにおけるステップ104では、定義された目的関数を最適化するような始点からの距離区間、方位区間が算出される。本実施の形態では、後述するように中間テーブルを作成して高速化処理が図られている。この中間テーブルでは、始点集合を母点とする幾何図形であるボロノイ図(ティーセン分割)を、逐次添加法によって作成している。即ち、 m 個の母点 P_1, \dots, P_m からなるボロノイ図 V に新たな母点 P_{m+1} を加え、 $m+1$ 個の点からなるボロノイ図 V

v_i を作っている。

【0028】図8は、逐次添加法による処理の流れを示した図である。図の左側は処理の流れをフローチャートで示し、図の右側に形成過程におけるボロノイ図の例を示している。まず、 P_1, \dots, P_n の中で P_{m_i} に最も近い点 P を「ボロノイ図の高速点位置決定法」(後述)により求める(ステップ111)。次に、 P のボロノイ領域内に P_{m_i} の垂直2等分線 L を引く(ステップ112)。次に、 L が接するボロノイ領域の母点に対しても同様に垂直2等分線を引く(ステップ113)。上記の作業を繰り返し、 P_{m_i} のボロノイ領域を作成して V_{m_i} とする(ステップ114)。

【0029】図9は、ボロノイ図の点位置決定法における処理の流れを示した図である。ここで示す「ボロノイ図の点位置決定法」は、ある点 P が m 個の母点 P_1, \dots, P_m からなるボロノイ図 V のどの母点に最も近いかを求める(どのボロノイ領域にあるかを求める)ものである。まず、 P_1, \dots, P_m の中で任意の点 P_i を選び、点 P との距離 d を求める(ステップ121)。次に、 P_i の各隣接母点 P_j と P との距離 d_j を計算し、 d と比較する(ステップ122)。ここで、 $d_j < d$ が判断され(ステップ123)、 $d_j < d$ であれば $d = d_j$ 、 $P_i = P_j$ としてステップ122に戻る。そのような母点がない場合には、 P_i を答えて(ステップ125)終了する。

【0030】図10は、逐次添加法の前処理における処理の流れを示した図である。まず、母点が各ピクセルにおよそ1つ含まれるような深さ d の四分木を作成する(ステップ131)。次に、図(a)のようにピクセルに番号を付ける(ステップ132)。この図(a)は、地図(2次元平面)のビューを示しており、ここに示される矢印は、ピクセル番号を付ける順序を示したものである。次に、各母点を座標値によってそれぞれピクセル(四分木の葉)に割り当て、各葉にラベルを付ける(ステップ133)。次に、番号順に、各葉の全ての(まだラベルのない)先祖に対して、自分のラベル値をコピーする(ステップ134)。最後に、この四分木の各節と葉を、広さ優先順に並べ、その順に「逐次添加法」を実行する(ステップ135)。図(b)は深さ3の四分木のビューを示しており、深さ方向ではなく横方向の順番にて「逐次添加法」が実行される。尚、駅、郵便局、警察、学校、公園など、分析対象になり易く、あまり位置的に変更のない地図上のエンティティについては、予めボロノイ図を構成しておくこともできる。

【0031】その後、ボロノイ図を利用した始点からの距離が計算される。ここでは、「 n 個の質問点に対し、 m 個の母点 P_1, \dots, P_m からなるボロノイ図 V を使った点位置決定問題を解く」ことが行なわれる。即ち、 n 個(かなり多い数)の各質問点(例えば犯罪データなど)と、 m 点(コンビニなど)からなる始点集合との距離が計算される。但し、距離は、最寄の始点との距離とする。より

具体的には、ピクセルで最初の質問点に対しては、逐次点においてそのピクセルのラベルとなった母点により、質問点が各ピクセル毎に計算される。また、ピクセルで最初の質問点でない場合には、前回の距離・方位計算で最近点とされた母点によって質問点が各ピクセル毎に計算される。以上のようにして、中間テーブルが作成される。

【0032】最後に、距離、方位のバケット集計と最適距離計算がなされる。ここでは、 n 個(この n はかなり多い数)の各質問点の属性値を、距離・方位の値に応じて集計する。即ち、各バケット毎に「データ数」と目的関数値算出に必要なデータが集計される。例えば、「顧客スキーマ」(支持率 S 以上の顧客の「平均年収」最大化距離)という目的関数の場合には、距離に応じたデータ数と年収合計数が順次、加算され、集計結果が出力される。このような集計情報を1回スワイプすることで、最適化距離が求められる。この結果は、例えば図1や図2に示したように、地図上に円形状や扇形状等によって表示される。

【0033】以上、本発明における処理のアルゴリズムを説明した。このような処理アルゴリズムは、コンピュータ・プログラムによって実現し、実行することができる。図11は、空間データマイニング装置としてのコンピュータ・システムの概略構成を説明するための図である。本実施の形態における処理アルゴリズムは、図示するようなコンピュータ・システムにおいて実行可能なプログラムとすることもできる。処理プログラムは、ハードディスクドライブ(HDD)75に格納され、実行時にはメインメモリ72にロードされ、CPU71によって処理される。また、HDD75は住所などの空間情報を含む大量のデータベースをも含んでおり、処理プログラムはそのデータベースに対するアクセスを行うものである。地図情報システム(GIS)等の地図情報や最適化距離、最適化方位の計算結果は、表示装置76によってユーザに提示される。ユーザは、入力装置77にて調べたい目的関数の入力や、データ出力の命令等を入力する。このような入力装置77には、キーボードやマウス、ポインティング・デバイスやディジタイザ等を含む。さらに、出力結果を補助記憶装置であるフロッピー(登録商標)ディスクドライブ(FDD)73のフロッピーディスクに記憶したり、また新たなデータをFDD73から入力することもできる。さらに、CD-ROMドライブ74を用いて、データを入力することもできる。

【0034】更に、本実施の形態における処理アルゴリズムを実現したコンピュータ・プログラムは、フロッピーディスクやCD-ROMといった記憶媒体に記憶して、持ち運ぶことができる。この場合、通常のデータベース検索プログラムのデータ取り出し部分や、表示装置76に表示するだけの処理を行うプログラムは、すでにHDD75に記憶されている場合もある。従って、それ

以外の部分が上記のような各種記憶媒体にて流通することは通常行われる事項である。また、図示されていない通信装置がバス78に接続されて、遠隔地にあるデータベースを用いて処理したり、処理結果を遠隔地に送信するように構成することも可能である。即ち、住所などの空間情報を含む大量のデータベースを図11に示す構成の外部に設けるように構成することもできる。

【0035】次に、本実施の形態における構成を、更に、機能ブロック図等を用いて詳述する。図12は、本実施の形態における空間データマイニングシステムの構成を説明するためのブロック図であり、図11で示したCPU71における構成を詳述したものである。ここでは、大きく、中間テーブル作成部30と最適距離計算部39とを備えている。この中間テーブル作成部30は、ボロノイ図作成部31、距離計算部32、および距離別計算部33とを備えている。このボロノイ図作成部31は、ID、名前、地図上の座標等からなる、例えばコンビニの点集合等の始点集合データを受けて、始点添加順決定部34およびボロノイ図点追加部35によりボロノイ図を作成している。距離計算部32では、ID、名前、地図上の座標、買上金額等からなる、例えば顧客データの点集合等の質間点集合データを受けて、距離の計算された顧客データレコードあるいはレコード集合を生成している。また、距離別計算部33では、距離計算部32から出力された顧客データレコードを基に、距離区間毎に最適化に必要なレコード値を集計している。その結果として、図12に示されるような中間テーブルが作成される。ここでは、距離区間毎にレコード値と買上金額計が示されている。最適距離計算部39では、この中間テーブルをスキャンして最適化関数値の最も良くなる距離が計算され、距離最適化ルールとして出力がなされる。

【0036】次に、図13および図14を用いてボロノイ図作成部31における始点添加順決定部34の機能を説明する。図13は、ボロノイ図作成部31における始点添加順決定部34の構成を説明するためのブロック図である。この始点添加順決定部34は、平面四分割機能41、始点分配機能42、代表点決定機能43、四分木作成機能44、および添加順序決定機能45を備えている。平面四分割機能41は、始点集合データを全て入力し、入力始点数 m に応じて、図に示すように平面四分割を $t(t=m^{1/2}-1)$ 回、繰り返す。即ち、各段の1ピクセルを4等分するが、各ピクセルには、ほぼ1点の始点を有することが好ましい。

【0037】図14は、分割された平面ピクセルと作成された四分木を説明する図である。図14に示すように、平面四分割機能41によって分割されたピクセルには番号が付けられる。ここでは深さ3として0~63の64のピクセルに分割される。図13に示す始点分配機能42では、 m 個の入力始点が末端の分割平面(ここで

は0~64)のピクセルに振り分けられる。図14では、ピクセル0に始点8(S8)と始点19(S19)が、ピクセル60に始点12(S12)が振り分けられている。次に、図13に示す代表点決定機能43では、各ピクセルの中にある始点の中から、一つを、そのピクセルの代表点として選定する。例えば、図14のピクセル60では始点12(S12)が選ばれている。複数始点がある場合には、任意の一つを選び(例えば、ピクセル0では始点8(S8)が選ばれている)、始点がない場合には、代表点は「なし」となる。

【0038】図13に示す四分木作成機能44では、前述の四分割の繰り返し過程で現われる、(最下段ではない)各段の各ピクセルを中間ノードとする四分木が作成される(図14右図参照)。更に、木の下段の中間ノードから順に、各中間ノードの代表点が決定される。図14では、各中間ノードにおける子ノードの代表点の中から一つを代表点としており、始点8(S8)が順に選ばれている。図13に示す添加順序決定機能45では、四分木の各ノードを広さ優先順に最上段から走査していく。即ち、深さ方向(深さ優先順)ではなく、各段ごとに横方向に走査する。その過程で、各ノードのまだリストにない代表点を出力始点リストの最後に加えていく。最下段の葉ノードに関しては、代表点だけではなく、そのノードに属する全ての(まだリストにない)始点に加えらる。即ち、図14の例では、ピクセル0の始点19(S19)が加えられる。以上のようにして、始点添加順決定部34から、順序付けされた始点集合が図12に示したボロノイ図点追加部35に出力される。

【0039】図15は、図12に示したボロノイ図作成部31のボロノイ図点追加部35の構成を説明するためのブロック図である。このボロノイ図点追加部35は、大きく高速点位置決定機能51と領域分割機能52を有している。更にこの高速点位置決定機能51は、開始位置決定機能53および漸近機能54とを備え、 P_1, \dots, P_n の中で P_{pi} に最も近い点 P を「高速点位置決定法」により求めている。この開始位置決定機能53では、入力点が四分木の最下段でどのピクセル X に属するかが計算され、ピクセル X の代表点の始点(ボロノイ点) P_i を開始位置とされる。但し、代表点「なし」の場合には、親ノードの代表点を開始位置とされる。また、漸近機能54では、 P_i と入力点 P_{pi} との距離 d が計算される。そして、 P_i の各隣接ボロノイ点 P_j と P_{pi} との距離 d_j が計算され、 d との比較がなされる。 $d_j < d$ ならば $d = d_j$ 、 $P_i = P_j$ とされる。そのような隣接母点がなければ、 P_i が答えとされる。そして、ピクセル X の代表点は P_i とされる。領域分割機能52では、右図に示すように、 P のボロノイ領域内に PP_{pi} の垂直2等分線 L がひかれる。 L が接するボロノイ領域の始点を P とし、同様に垂直2等分線がひかれる。これを垂直2等分線で囲まれる領域ができるまで繰り返されることによ

て、右図のような $m+1$ 点からなる中間ボロノイ図 V_m を得ることができる。

【0040】図12に示した中間テーブル作成部30では、以上のようにして作成されたボロノイ図作成部31にて作成されたボロノイ図と、質問点集合データから、前述したように、距離計算部32によって距離の計算された顧客データレコードが出力される。即ち、質問点集合の各レコードにおける質問点レコードと中間ボロノイ図 V_m とから、高速点位置決定法によって入力点に最も近いボロノイ点が抽出され、質問点レコードとボロノイ点との距離を計算し、距離付き質問点レコードを出力する。その後、前述したように、距離別計算部33を経て、距離区間毎にレコード値が集計された中間テーブルが中間テーブル作成部30から出力される。出力された中間テーブルの情報は、例えばHDD75等に格納される。

【0041】このソートされた中間テーブルを用いて、最適距離計算部39にて最適距離が計算される。例えば、図12に示したような質問点集合データにおける顧客データであれば、距離の小さなレコードから順に、買上金額をアキュムレート(累算)し、その過程で「アキュムレートされた買上金額/これまでのアキュムレートされたレコード数」の最も高い値を出す距離を記録する。その際、一般的には、このアキュムレートの過程で、最高値を出したレコードと、中間テーブルにおけるその次のレコードの距離との中間値を暫定最適距離として記録する。このように目的関数の最高値を出す暫定最適距離を維持しながら中間テーブルを1度スキャンすることで、最適距離を求めることができる。求めた最適距離は、表示装置76を用いて図1のように表示することが可能である。

【0042】方位最適化を図るアルゴリズムの中間テーブルには、図2に示したように最適化する対象である各円の内側のデータのみが含まれる。即ち、空間情報と関連付けられているデータベースの各レコード(質問点)に対して、基準点との距離を計算し、半径以内に含まれているものについて基準点との角度を計算し、角度の値に応じて中間テーブルのレコードへ集計される。但し、検索対象となる円が互いに重なっている場合には、複数の円に同時に含まれる質問点が存在する場合があり得ることから、中間テーブルを作成する際に、以下の二通りのどちらかの方法で処理を行うこととする。

・それぞれの基準点に基づく角度を、全ての対応する中間テーブルのレコードに加える。

*

距離 ユークリッド距離、

始点集合の定義 郵便局(ALL, 各種類)、学校(ALL, 各種類)、

目的関数の定義 顧客(支持率S以上の顧客の平均年収最大化距離)

顧客(性別の相互情報量最大化距離)

を指定した場合、各始点集合から各目的関数の最適化ルールは、例えば、

50

* 一番距離の近い基準点に対する角度に対応するレコードにのみ、中間テーブルへ加える。

【0043】方位最適化の際の中間テーブルには、このように等間隔の昇順あるいは降順の角度区間を属性として有している。例えば、目的関数が平均年収である場合に、中間テーブルは、その計算に必要な年収合計とレコード数とを属性として持つ。

年収合計	角度区間	レコード数
8,400,000	0~10	6
5,000,000	10~20	1
7,000,000	20~30	3
5,800,000	30~40	3

【0044】その後、中間テーブルを用いて目的関数の最適方位の範囲を決定する。例えば、角度の小さなレコード順のインデックスを s, t として、区間 $[s, t]$ の平均年収 $X[s, t]$ を、 $X[s, t] = \text{「区間}[s, t]\text{」の間に含まれるレコードの年収の総和} \div \text{区間}[s, t]\text{の給与所得者数} \text{」}$ と表わすと、 $X[s, t]$ を最適化する区間 $[s, t]$ を求める問題となる。そのための区間アルゴリズムは、中間テーブルサイズを n として $O(n)$ で実行可能であるが、角度を数値化した時に発生する0度での不連続性を考慮したアルゴリズムを適用する。例えば、ソートされた中間テーブルのレコードに対し、年収とレコード数とをそれぞれ累積演算させながら、

・ Σ 年収 / Σ レコード数を最大にするレコード位置 t と t における Σ 年収と Σ レコード数

・ Σ 年収 / Σ レコード数を最小にするレコード位置 s と s における Σ 年収と Σ レコード数

を記憶しておく。360度までスキャンしたときの値から、最適方位を求めることが可能である。 $s < t$ ならば、最適区間は0をまたがずに $[s, t]$ と与えられ、その平均年収も直ぐに求めることができる。 $t < s$ ならば、最適区間は0をまたぐ $[s, t]$ で、その平均年収も全質問点数と全質問点との年収総額を用いて求めることが可能である。このようにして求めた結果は、表示装置76を用いて図2のように地図上に表示することが可能である。

【0045】このように、本実施の形態によれば、必要に応じて入力パラメータとして距離の定義(ユークリッドか(この場合ほぼ線形時間)ネットワークか(この場合多項式時間))や方位(方角)の定義、始点集合の定義、および目的関数の定義を指定することで、最適化ルールを計算して列挙することが可能となる。例えば、

「平均年収を最大化するのは郵便局からX以内(支持率 s 、平均年収 x)」

「性別の相互情報量を最大化するのは大学からX以内(支持率s, エントロピーゲインg)」

などのようなルールがそれぞれの組み合わせに対して列挙され、ユーザは自分の興味に従って、これらのルールをソートしたり、ファイリングしたりすることが可能となる。また、特に関心があるものについては、図1や図2に示したようにGUIを用いて地図上にルールを表示することもできる。

【0046】

【発明の効果】以上説明したように、本発明によれば、予め距離や方位を定めた上で空間相関ルールを導き出すのではなく、多くの分析業務にて要求される、ある目的を最適化する距離自体、方位自体を求める空間データマイニングを提供することが可能となる。

【図面の簡単な説明】

【図1】 本実施の形態におけるモデリング1として、距離最適化エンジンによる出力例を示した図である。

【図2】 本実施の形態におけるモデリング2として、方位最適化エンジンによる出力例を示した図である。

【図3】 本実施の形態における距離・方位最適化エンジンのアルゴリズムの概略を説明するためのフローチャートである。

【図4】 データベース例を示した図である。

【図5】 入力パラメータとして始点群や基準群の定義例を示した図である。

【図6】 距離や方位の定義例を示した図である。

【図7】 目的関数の定義例を示した図である。

【図8】 逐次添加法による処理の流れを示した図である。

【図9】 ボロノイ図の点位置決定法における処理の流れ

*れを示した図である。

【図10】 逐次添加法の前処理における処理の流れを示した図である。

【図11】 空間データマイニング装置としてのコンピュータ・システムの概略構成を説明するための図である。

【図12】 本実施の形態における空間データマイニングシステムの構成を説明するためのブロック図である。

【図13】 ボロノイ図作成部31における始点添加順決定部34の構成を説明するためのブロック図である。

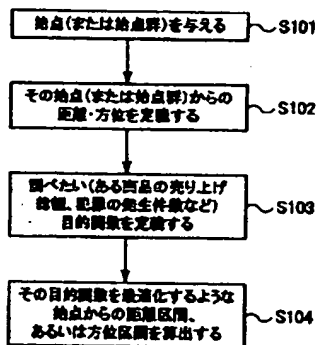
【図14】 分割された平面ピクセルと作成された四分木を説明する図である。

【図15】 図12に示したボロノイ図作成部31のボロノイ図点追加部35の構成を説明するためのブロック図である。

【符号の説明】

11…地図、12～14…始点(または始点群)、15～17…円、18～20…方位区間、30…中間テーブル作成部、31…ボロノイ図作成部、32…距離計算部、33…距離別計算部、34…始点添加順決定部、35…ボロノイ図点追加部、39…最適距離計算部、41…平面四分分割機能、42…始点分配機能、43…代表点決定機能、44…四分木作成機能、45…添加順序決定機能、51…高速点位置決定機能、52…領域分割機能、53…開始位置決定機能、54…漸近機能、71…CPU、72…メインメモリ、73…フロッピーディスクドライブ(FDD)、74…CD-ROMドライブ、75…ハードディスクドライブ(HDD)、76…表示装置、77…入力装置、78…バス

【図3】



【図4】

データベース例

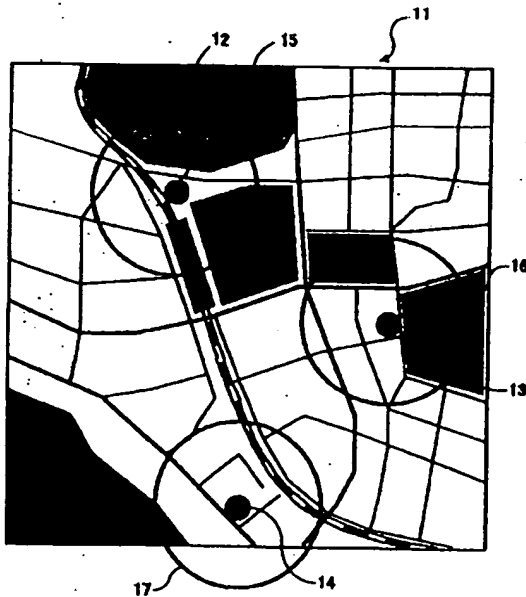
郵便局スキーマ=(ID, 位置(座標), 種類)
 学校スキーマ=(ID, 位置(座標), 種類)
 警察スキーマ=(ID, 位置(座標), 種類)
 駅スキーマ=(ID, 位置(座標), 乗降客数, 乗降駅名)
 コンビニスキーマ=(ID, 位置(座標), 売上, 店舗名)
 顧客スキーマ=(ID, 位置(座標), 年齢, 性別, 年収, 職業)
 ATMスキーマ=(ID, 位置(座標), 平均出金額, 平均利用回数)
 犯罪スキーマ=(ID, 位置(座標), 種類)

——は数値属性
 ——はカテゴリー属性

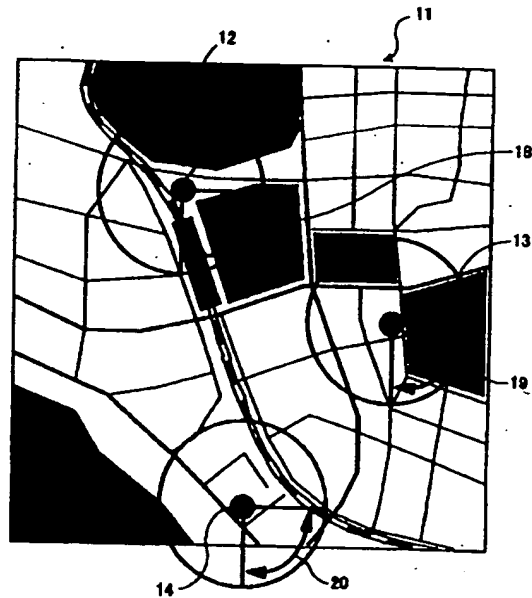
(12)

特開2001-318938

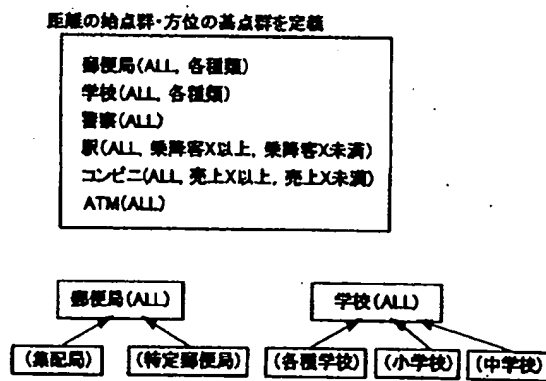
【図1】



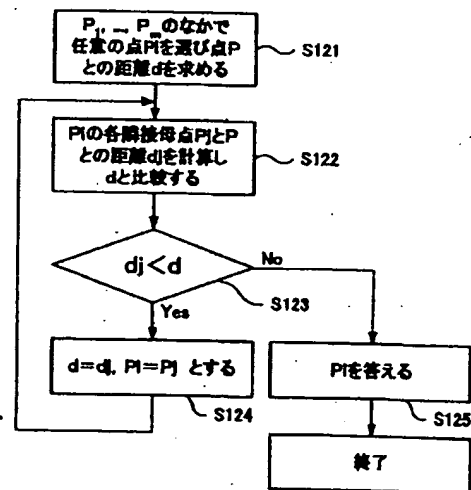
【図2】



【図5】



【図9】

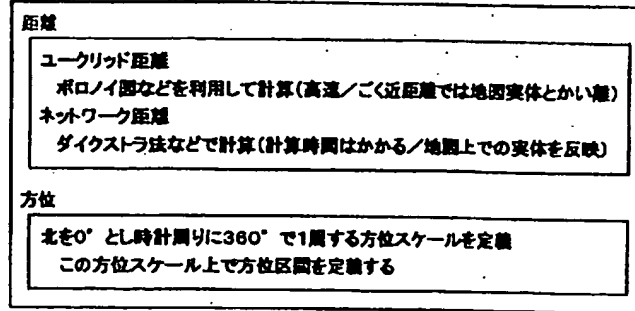


(13)

特開2001-318938

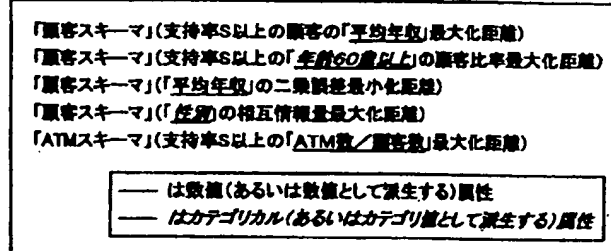
【図6】

距離の定義・方位の定義

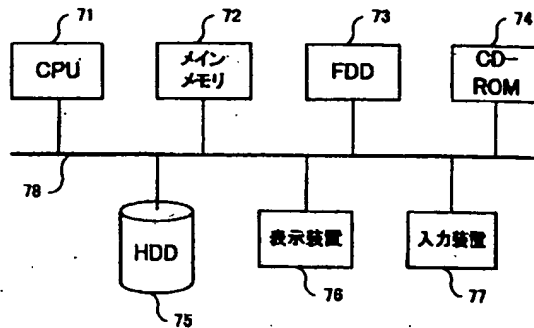


【図7】

目的関数の定義



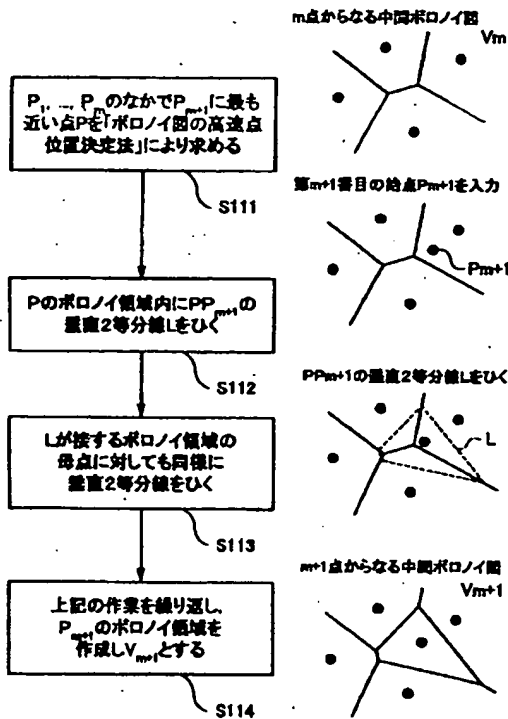
【図11】



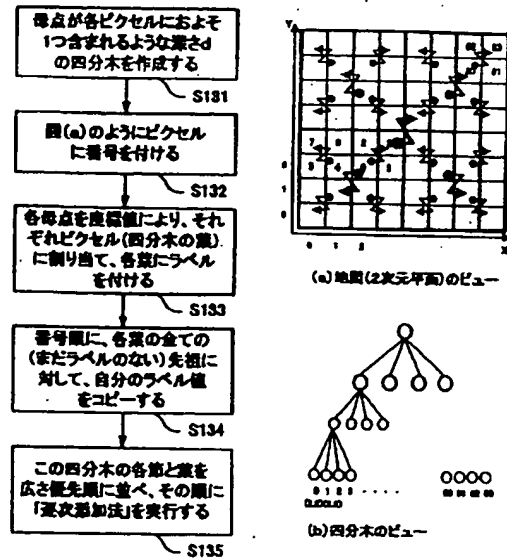
(14)

特開2001-318938

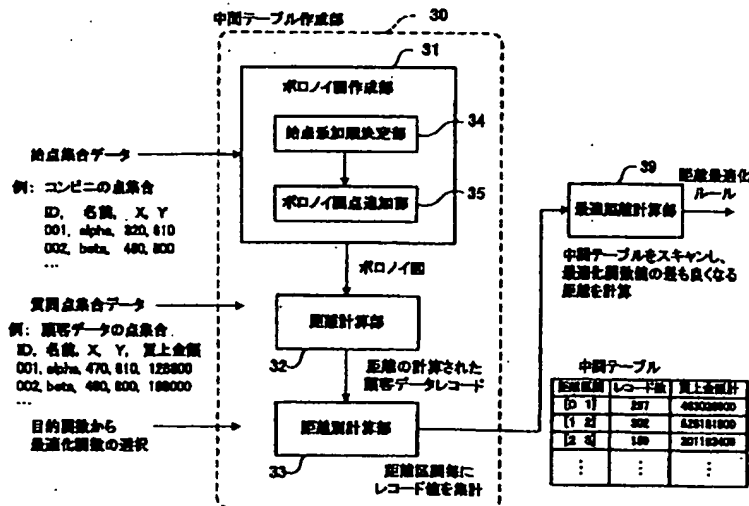
【図8】



【図10】



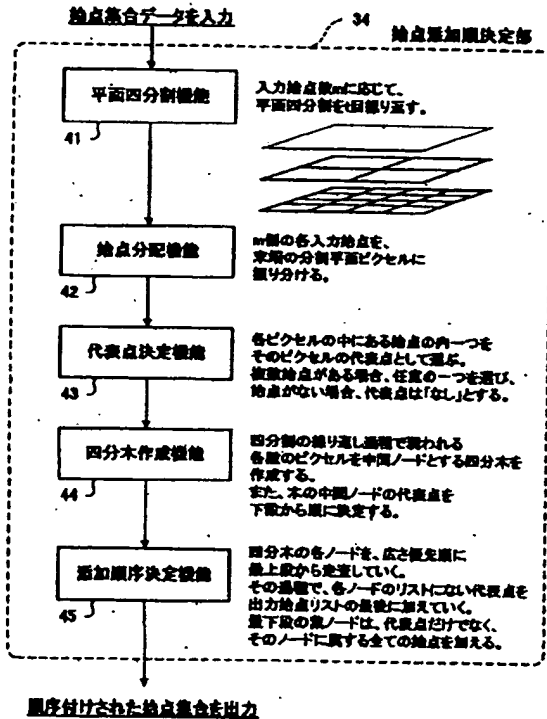
【図12】



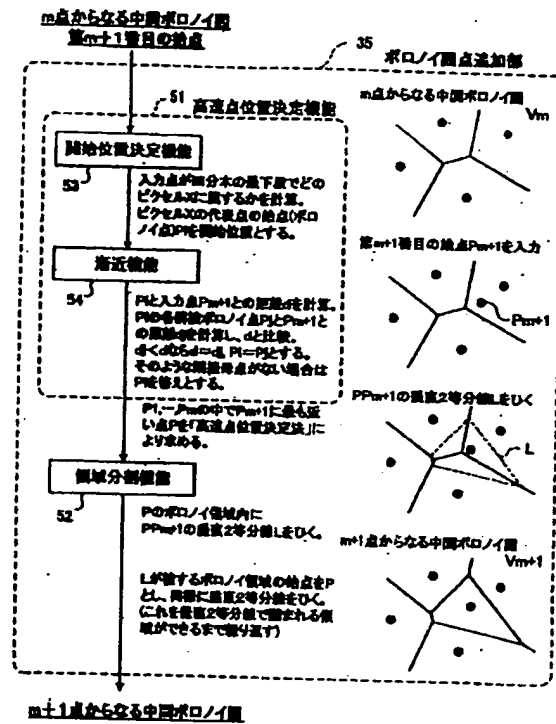
(15)

特開2001-318938

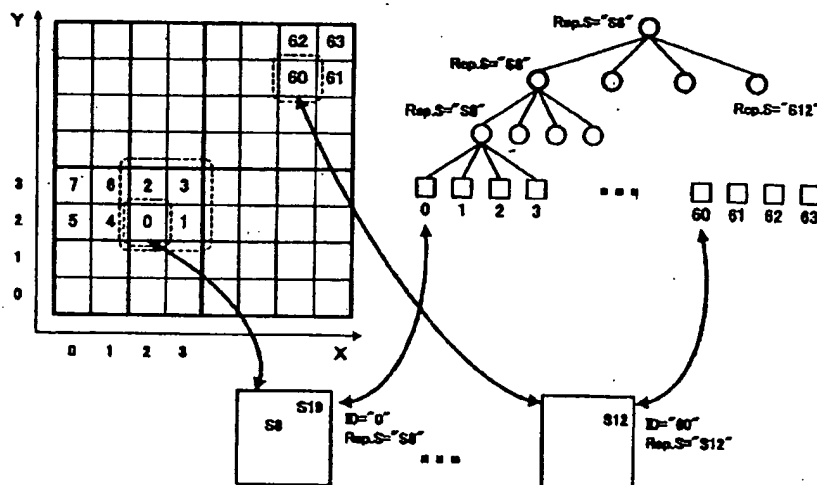
【図13】



【図15】



【図14】



BEST AVAILABLE COPY

(16)

特開2001-318938

フロントページの続き

(72)発明者 森本 康彦
神奈川県大和市下鶴間1623番地14 日本ア
イ・ビー・エム株式会社 東京基礎研究所
内
(72)発明者 久保 晴信
神奈川県大和市下鶴間1623番地14 日本ア
イ・ビー・エム株式会社 東京基礎研究所
内

Fターム(参考) 2C032 HB05
5B049 AA06 EE02 EE31 FF01
5B075 PP02 PP03 PQ02 PQ13 PQ69
UU14
5L096 BA12 DA02 FA66 FA67-FA74
GA34 MA07
9A001 BB06 FF01 GG04 JJ01 JJ11